Understanding Enterprise Architecture with Topic Modeling

Preliminary Research Based on Journal Articles

Nardello, Marco; Møller, Charles; Gøtze, John

Published in:
ICEIS 2018 - Proceedings of the 20th International Conference on Enterprise Information Systems

DOI (link to publication from Publisher):
10.5220/0006693106400648

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Understanding Enterprise Architecture with Topic Modeling
Preliminary Research based on Journal Articles

Marco Nardello[1,*], Charles Møller[1] and John Gøtze[2,3]
[1]*Department of Materials and Production, Aalborg University, Aalborg, Denmark*
[2]*IT University of Copenhagen, Copenhagen, Denmark*
[3]*QualiWare aps, Farum, Denmark*

Keywords: Enterprise Architecture, Topic Modelling, Content Analysis, Research Theme, Future Research, Machine Learning, Latent Dirichlet Allocation.

Abstract: The next 3 years will be more important than the last 50 due to the digital transformation across industries. Enterprise Architecture (EA), the discipline that should lead enterprise responses to disruptive forces, is far from ready to drive the next wave of change. The state of the art in the discipline is not clear and the understanding among researchers and practitioners is not aligned. To address these problems, we developed a topic model to help structure the field and enable EA to evolve coherently. In this preliminary study, we present the 360 identified topics in EA literature and their evolution over time. Our study supports and combines the findings from previous research and provides both a deeper analysis and more detailed findings.

1 INTRODUCTION

Enterprise Architecture (EA) leads enterprise responses to disruptive forces (Gartner Inc. 2017) and industries are now on the brink of digital transformation. Globally 72 percent of CEOs believes that "the next 3 years will be more critical for their industry than the last 50 years" (KPMG 2016, p.8). KPMG also explained that "the speed of change will be exponential" and 77 percent of CEOs "are concerned about whether their organization is keeping up with new technologies" (KPMG 2016, p.7). However, the current state of the art of EA is not clear, and its fundamental concepts are not shared among researchers and practitioners (Saint-Louis & Lapalme 2016; Rahimi et al. 2017). In addition, from informal interviews the authors acknowledged that EA researchers and practitioners find problematic to position themselves in the field because they are lacking its overview.

Two studies tried to address this problem systematically (Simon et al. 2013; Saint-Louis & Lapalme 2016) though their limited scope left important research themes uncovered. Simon et al. (2013) conducted a content analysis of EA including articles published until 2010. Since then EA literature has almost tripled (Saint-Louis 2016). A more recent example, the Systematic Mapping Study (SMS) of Saint-Louis and Lapalme (2016) focused on literature from selected journals therefore excluding most literature on EA. In addition, based on Debortoli's et al. (2016) article both studies are subject to human bias. In the first, the authors conceptualized EA in a model and classified literature applying it (Debortoli et al. 2016). In the second one, the authors manually identified topics.

These studies exclusively identified the topics without specifying the concepts related to them and their evolution over time. To date there is no study that identifies topics in EA without major human subjectivity bias. In addition, directions for future research are mainly based on personal experience (Korhonen et al. 2016; Lapalme et al. 2016).
Another problem, not restricted to EA, is the lack of a structure of research fields and how they evolved over time.

To identify EA's topics and their evolution over time we developed a Topic Model applying Latent Dirichlet Allocation (LDA) method by Blei et al. (2003). This method develops a list of topics, for each one presents the most frequently occurring words and identifies the most relevant literature (Blei et al. 2003). The LDA method has been applied successfully in other fields like in Information Systems research (Chen & Zhao 2015). In this preliminary study, we investigate the following

Nardello, M., Møller, C. and Gøtze, J.
Understanding Enterprise Architecture with Topic Modeling.
DOI: 10.5220/0006693106400648
In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)*, pages 640-648
ISBN: 978-989-758-298-1

research questions:

1. What are the topics in EA literature?
2. How did topics in EA literature evolve over time?

We present a topic model of EA based on 602 journal articles. An article following this preliminary study will present a topic model including also non-journal publications.

This study aims to guide coherently the development of EA research and it mainly contributes to the EA community in two ways. First, it allows the community to identify topics in EA that are relevant for them, opening new research opportunities. Second, it identifies the literature related to each topic.

A final contribution to research in general is to present a new approach for structuring research fields through the development of a topic model. This type of efforts can support systematic literature reviews by structuring the field and providing the context and literature for this type of research.

The article continues presenting the background literature in section 2, the method applied in section 3 and the key findings of this study in section 4. We discuss the results in section 5 and the conclusions and future research are presented in section 6.

2 BACKGROUND

The foundations of EA date back to the 1960s when IBM initiated the Business Systems Planning (BSP) methodology (Kotusev 2016). BSP was “a structured approach to assist a business in establishing an information systems plan to satisfy its near- and long-term information needs” (IBM 1978). In the 1980s, EA emerged from BSP (Kotusev 2016). Zachman (1987) developed a framework for information systems and Spewak and Hill (1993) defined Enterprise Architecture Planning. EA Planning was “the process of defining architectures for the use of information in support of the business and the plan for implementing those architectures” (Spewak & Hill 1993). Finally at the beginning of the new millennium, The Open Group Architecture Framework (Group 2009) developed a new EA reference architecture and methodology that are today widely used in practice (Simon et al. 2013).

Two articles mapped EA contributions (Simon et al. 2013; Saint-Louis & Lapalme 2016). The first one is a combined bibliometric study and content analysis of EA research from Simon et al. (2013). In their study, the top down content analysis classified EA publications in a predefined scheme. In this way, the authors biased the results of the content analysis by defining the classification scheme based on their understanding prior the study (Debortoli et al. 2016, pp.112–113). In addition, analysing Saint-Louis' and Lapalme's EA article distribution by year (2013, p.76) is possible to infer that Simon et al. (2013) study did not consider almost two thirds of EA literature currently available. This is also supported by a search on Scopus database for articles with "Enterprise Architecture" in title, abstract and keywords. Around 1300 articles were published before 2010 and almost 2000 were published after.

The study of Saint-Louis and Lapalme (2016) reviewed 171 journal articles. They presented a bibliometric study and a summary of ten conceptualizations of EA, namely framework, model, discipline, integration, measurement, strategy, principles, design, literature, and practitioner. One limitation of their study is to have included only journal articles from few sources. Other sources were not considered. In addition, when the authors classified EA topics they assigned only one topic per article. This is restrictive because it is common for a research article to include multiple topics (e.g., an article about EA models for business alignment covers both the EA model topic and business alignment topic) (Chen & Zhao 2015, p.2).

Another challenge in EA research is that efforts are weakly integrated (Simon et al. 2013, p.19). This is supported by the fact that systematic studies in EA are very limited in scope. Excluding the SMS previously mentioned, the other three available in the field focus respectively on enterprise integration (Banaeianjahromi & Smolander 2016), applications of ontologies (Pinto et al. 2014), measurement and indicators (Abdallah et al. 2016). Among the systematic literature reviews, we found areas as the automated analysis and documentation (Florez et al. 2016; Farwick et al. 2016; Hauder et al. 2012), aspects of EA Management (Kotusev 2017; Rahimi et al. 2017; Jugel et al. 2016; Huber et al. 2017; Brosius 2016; Lange et al. 2016; Schneider et al. 2013; Wißotzki & Sonnenberger 2012) and a research group involved in EA implementation (Rouhani et al. 2015; Nikpay, R. Ahmad, et al. 2017; Nikpay, R. B. Ahmad, et al. 2017).

Our study aims at overcoming the limitations of previous research, outline the topics in EA research, and present their evolution over time.

3 METHODOLOGY

There are five main methods for performing text-categorization (Debortoli et al. 2016) – bottom-up and top-down manual coding, dictionaries, and supervised and unsupervised machine learning. We applied the latter that uses documents to inductively discover categories and assigns documents to these categories (Debortoli et al. 2016). We chose this method because it "generates reproducible results that are not subject to the human subjectivity bias" (Debortoli et al. 2016).

An application of unsupervised machine learning is probabilistic topic modelling. This approach is based on the assumption that "words that occur in the same contexts tend to have similar meanings" (e.g. the co-occurring words "mozzarella", "tomato", "basil", "margherita", "oven" all refer to the "pizza" topic) (Turney & Pantel 2010, p.142). Three main methods are used in Information System research (Debortoli et al. 2016): Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA) and Leximancer. We applied LDA because previous studies proved that LSA methods suffer from interpretability issues (Debortoli et al. 2016) and Leximancer algorithms are scarcely documented (Debortoli et al. 2016).

LDA methods are grounded on the assumption that documents are generated from a set of topics and that each topic is characterized by a distribution over words (Chen & Zhao 2015, p.2). For this reason, a document can contain multiple topics. In this way, we identified research topics in EA discipline.

In this study, for the data gathering and data filtering steps we followed (Petersen et al. 2015), and for the data preparation, application of text-mining techniques, and evaluation of the topic model we followed (Boyd-Graber et al. 2014; Debortoli et al. 2016). The first three steps have been performed manually while the last two were automated.

3.1 Data Gathering

We included journal articles that contained the words "Enterprise Architecture" in the title, abstract or list of keywords. The search was performed in July 2017 on the following databases (in parenthesis the number of articles retrieved): Scopus (539 articles), ABI/INFORM (167 articles), Business Source Premier of EBSCO (205 articles), Web of Science (287 articles), Compendex (306 articles), INSPEC (0 articles), IEEE Xplore Digital Library (31 articles) and AIS Electronic Library (61 articles). The search generated 715 unique articles. Depending on the database the search was restricted to journal articles or to peer-reviewed articles. When possible, we limited our search to articles written in English. We decided to include only journal articles to focus this preliminary study on what we consider the most reliable outlet of contributions in the field.

3.2 Data Filtering

Each combination of title and abstract represented what we will refer to as a document. The first author performed this filtering alone since in case of a large number of studies and many of them are clearly identifiable noise the process may be conducted individually (Petticrew & Roberts 2008). When in doubt, he was inclusive. The following criteria were applied based on (Petersen et al. 2015). Inclusion criteria:

- The study is published in a journal
- The study relates to EA, as defined in the Introduction section.

Exclusion criteria:

- Conference or journal editorials
- Studies presenting book reviews
- Studies not written in English
- Books and grey literature

The filtering process had two steps, see Figure 1. First, we removed all the articles that were not journal publications. Second, we removed articles not related to EA based on title and abstract. We used the resulting 602 articles as input for the topic model.

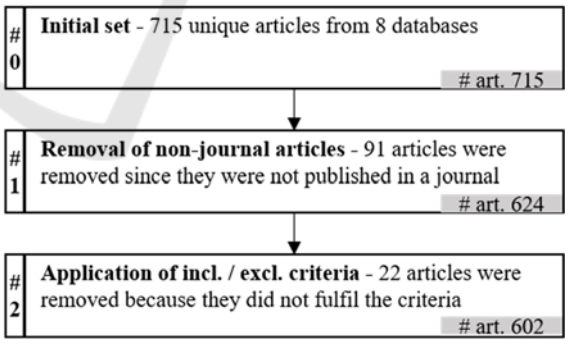

Figure 1. Filtering process.

3.3 Data Preparation

When preparing the documents for the analysis we followed the guidelines from Boyd-Graber et al. (2014). We started by removing punctuation notations from the documents (except for the dash that connects words closely related) and formatted the text in small caps. We did not apply stemming since

it combines terms that may have different meanings. Next, we used Blei's and Lafferty's Turtotopics to identify multi-word expressions (Blei & Lafferty 2009). They are also known as n-gram and "business IT alignment" is an example of 3-gram word that captures a single concept in EA. We used for this analysis only the abstracts with the following settings: p-value 0.1 and the n-word appears at least 3 times in the entire corpus. Since some phrases were repeated in shorter forms, we considered only the shortest form that preserved the meaning of the longer phrases (e.g. we considered "enterprise architecture framework" instead of "enterprise architecture framework eaf"). We modified the words in the document to distinguish them from 1-gram words (e.g. "EnterpriseArchitectureFramework").

Afterwards, we extended Stanford's Topic Modeling Toolkit list of English language stopwords (Ramage & Rosen 2011) with common words pertaining to EA literature (see the Appendix for the full list). We used this list to exclude all the non-value adding words from the analysis. Removing the *n* top words is an alternative approach, thought the result were worse than using the stopword list.

3.4 Text-mining

For the text-mining we used Stanford Topic Modeling Toolbox (v0.4.0) (Ramage & Rosen 2011). We computed the topic model with different number of topics 10-fold ranging from 10 to 450. The LDA method had the following parameters: max iteration 3000, topic smoothing 0.01, term smoothing 0.01. For training the model to fit the documents we experimented with the two main approaches – Gibbs sampler and variational Bayes approximation – and decided to use the latter since it produced better results. At 360 topics the results were most understandable. We did the topic labelling manually.

3.5 Evaluation

To evaluate the number of topics, we calculated the perplexity of held-out documents with different number of topics from 10 to 450 in 10-fold. "Perplexity is a standard performance measure of different models in natural language processing; a lower value of perplexity indicates a better model performance" (Chen & Zhao 2015). As the number of topic increases, the perplexity decreases until 360 topics and after that it stabilizes.

4 FINDINGS

In this study, we identified 360 topics in EA research, the keywords related to each of them and their evolution overtime. Due to space limitations, we will present only the ten topics that covered most of EA body of knowledge. For these topics, the five most recurring words and the percentage of coverage of the topic over the literature are available in the appendix. The full list of the topics is available online (QualiWare 2017).

First, the *Business Process* topic refers to business processes (e.g. in the financial industry) and includes concepts and information to support decision makers. Second, the *Framework* topic relates to EA frameworks that are adequate for integrating and structuring information. They can be used to capture the baseline structure of the organization as well as to clarify and achieve predetermined outcomes.

Third, the *Enterprise Resource Planning (ERP) Model* topic refers to models of companies ERP software specifying its requirements, features and functions. It is related to the *Business Process* topic and to the Business Process Reengineering practice.
Fourth, the *Management Activity* topic presents activities done by managers that include objectives, opportunities, clients, necessary resources, information, and information security. These activities relate also to the concept of holistic approach.

Fifth, the *Design* topic refers to the design of the components and their interaction of Business Processes, Systems, Information Technology (IT) in order to support the analysis and the collection of information.

Sixth, the *Modelling Methodology* topic refers to methodologies and approaches for modelling efficiently and systematically different facets of the organization. Two main fields of application are the healthcare sector and the Virtual Enterprise (VE).
Seventh, the *Analysis Method* topic includes the concepts of techniques, estimations, accuracy, uncertainty, and validation. This topic is scientifically investigated mostly through case studies.

Eight, the *Meta Model* topic is the one that covers the most body of knowledge. It related to meta-models and the *Analysis Method* and *Models* topics.
Ninth, the *Model* topic relates to integrated models that support the evolution of company's systems.

Finally, the *Service* topic refers to services and includes concepts like users, resources, infrastructure, and information systems.

For each topic, we plotted its usage over time by using Stanford Topic Modeling Toolbox feature

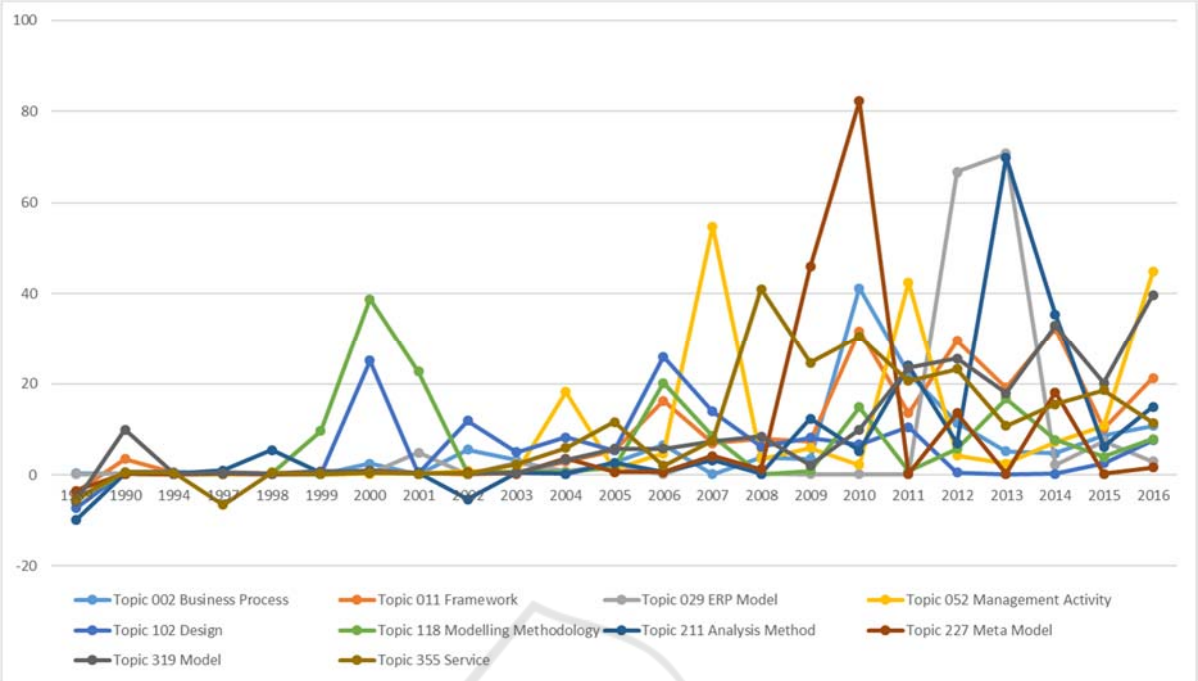

Figure 2. Topic distribution over time.

"slice". This feature creates subsets of the results associating it with one or more variables (e.g. year of publication, source, authors, and so on). By splitting the results for each year it is possible to understand when each topic emerged and which have a growing, stable or decreasing trend. Figure 2 presents the evolution of the topics previously described (the vertical axis represents the total number of words associated with each topic for each year).

Four topics emerged in the last century. The first topics that appeared in EA research are the *Framework, Modelling Methodology, Analysis Method* and *Model*. The *Framework* topic was introduced in 1990 and has it started being significantly researched from 2005 and it reached its peak in 2014. Research on the *Modelling Methodology* topic was popular around the year 2000 and since then has been less researched. The *Analysis Method* topic emerged in 1998 and it had a constant growth from 2008 until it reached its peak in 2013. Since then its has been less researched. Research on the *Model* topic started in 1990 and grew constantly from 2004 until now.

The remaining six topics were researched from the year 2000 onwards. Starting with the *Business Process* topic, it grew constantly from the year 2000 until 2010 and since then it became less researched. The *ERP model* topic emerged in 2001 and peaked in 2012 and 2013. Since then research on this topic decreased significantly. The *Management Activity* topic started to be more researched in 2004 and since then had a fluctuating trend. The *Design* topic emerged in the year 2000, it peaked in 2006 and since then had a decreasing importance. The *Meta Model* topic emerged in 2004 and peaked in 2010. Since then it had a fluctuating trend. Finally, the *Service* topic grew constantly from 2003 until 2008. Since then it had decreasingly researched.

Topic models also provide a new way of navigating literature. As an example of topic-based literature navigation we list the three articles that are contributing the most to the *Modelling Methodology* topic. First, Glazner (2011) applied simulation methods in EA. Second, Kim et al. (2006) developed a systematic modelling approach for VEs. Third, Nugraha et al. (2017) presented "a business architecture modeling methodology to support the integration of primary health care".

5 DISCUSSION

In this section, we will present the implications or our research and its limitations.

The three main implications of our study are to provide a better understanding of EA topics, integrate existing findings, identify new topics.

First, our study helps researchers and practitioners understand the topics in EA. It does so in three ways, identifying the concepts related to each topic, in which fields it has been mostly researched, and the literature related to it. For instance, the *Modelling Methodology* topic refers to modelling various facets of the organization and not exclusively the IT or business aspects. In addition, the focus of research has been on efficient and systematic approaches. Our topic model helps researchers and practitioners working with *Modelling Methodologies* to acknowledge of these concepts and the focus of research. Afterwards, the *Modelling Methodology* topic helps researchers and practitioners understand in which fields this topic is most applied, health care sector and VE. Finally, our study identifies the articles related to the topic.

Moving on to the comparison of EA topics with previous studies, we will use the top ten topics to illustrate how the topics identified in this study are related to previous studies. The details of this analysis are available in the appendix. Seven topics are covered in the content analysis of Simon et al. (2013) and six are related to the topics identified by Saint-Louis and Lapalme (2016). The topics in common are on a higher abstraction level than the others. The two that are not mentioned in previous studies are the *ERP Model* and *Service* topics. In this study, depending on the abstraction level of the topic some of them are inline and combine the findings of previous studies while others present new, and usually more specific, research topics.

This latter point combined with the topic-guided literature navigation can support literature studies by making available articles that might be left out from a keyword search.

Our study has three main limitations. First, the decision to include only journal articles might have caused significant contributions to be left out from this study. Having said so, this study is with Simon et al. (2013) among the ones that included the widest body of literature. Second, concerning the topics and their labelling, the model has been evaluated only by the authors. Even though both co-authors publish in the field of EA, external evaluation would improve the reliability of our results. A third limitation is to have analysed only the abstracts of the journal articles and not the full text. Having said so, this is a standard approach in Topic Modelling (Chen & Zhao 2015).

6 CONCLUSION

In this study, is a first step towards a map of topics of EA. This preliminary research identified topics in EA research based on 602 journal articles. We presented the ten topics that cover most literature in the field and based on these we discussed the contribution of topic modelling to EA research. We presented how these topics improve the understanding of EA literature, how they integrate and combine previous findings and how they can be used to navigate EA literature with a topic-guided approach. In addition, we have discussed how topic models can support systematic literature reviews in EA.

Future research will focus on extending the body of literature analysed to include also non-journal publications. In addition, different text mining approaches will be investigated to enhance the value of topic models in EA research. Interesting contributions that might be applied and further developed are the hierarchical topic modelling (Blei et al. 2010), correlated topic modelling (Blei & Lafferty 2007), automatic labelling technique (Lau et al. 2011) and supervised machine learning.

REFERENCES

Abdallah, A., Lapalme, J. & Abran, A., 2016. Enterprise Architecture Measurement: A Systematic Mapping Study. *2016 4th International Conference on Enterprise Systems (ES)*, pp.13–20. Available at: http://ieeexplore.ieee.org/document/7880427/.

Banaeianjahromi, N. & Smolander, K., 2016. What do we know about the role of enterprise architecture in enterprise integration? A systematic mapping study. *Journal of Enterprise Information Management*, 29(1).

Blei, D.M., Griffiths, T.L. & Jordan, M.I., 2010. The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM (JACM)*, 57(2), p.7.

Blei, D.M. & Lafferty, J.D., 2007. A Correlated Topic Model of Science. *The Annals of Applied Statistics*, 1(1), pp.17–35. Available at: http://www.jstor.org/stable/4537420.

Blei, D.M. & Lafferty, J.D., 2009. Visualizing topics with multi-word expressions. *arXiv preprint arXiv:0907.1013*.

Blei, D.M., Ng, A.Y. & Jordan, M.I., 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), pp.993–1022.

Boyd-Graber, J., Mimno, D. & Newman, D., 2014. Care and feeding of topic models: Problems, diagnostics, and improvements. *Handbook of Mixed Membership Models and Its Applications*, pp.225–254.

Brosius, M., 2016. Motivation for Coordination-a Complementary Approach to Enterprise Architecture Management Research. In *Enterprise Distributed Object Computing Workshop (EDOCW), 2016 IEEE 20th International*. IEEE, pp. 1–8.

Chen, H. & Zhao, J.L., 2015. ISTopic: Understanding information systems research through topic models. *Icis-Rp*, pp.1–10. Available at: http://aisel.aisnet.org/icis2015/proceedings/GeneralIS/6/.

Debortoli, S. et al., 2016. Text Mining For Information Systems Researchers: An Annotated Topic Modeling Tutorial. *Communications of the Association for Information Systems*, 39, pp.110–135. Available at: http://aisel.aisnet.org/cais%0Ahttp://aisel.aisnet.org/cais/vol39/iss1/7%0Ahttp://aisel.aisnet.org/cais%0Ahttp://aisel.aisnet.org/cais/vol39/iss1/7.

Farwick, M. et al., 2016. A situational method for semi-automated Enterprise Architecture Documentation. *Software and Systems Modeling*, 15(2).

Florez, H., Sánchez, M. & Villalobos, J., 2016. A catalog of automated analysis methods for enterprise models. *SpringerPlus*, 5(1). Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84962374165&doi=10.1186%2Fs40064-016-2032-9&partnerID=40&md5=9cbe5683010939375dbd92417b68553b.

Gartner Inc., 2017. Enterprise Architecture Definition. *https://www.gartner.com*. Available at: https://www.gartner.com/it-glossary/enterprise-architecture-ea/.

Glazner, C.G., 2011. Enterprise transformation using a simulation of enterprise architecture. *Journal of Enterprise Transformation*, 1(3), pp.231–260. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84971449078&doi=10.1080%2F19488289.2011.601398&partnerID=40&md5=374ea2c633d670d78b138ff93408fbc6.

Group, T.O., 2009. *The Open Group Architecture Framework TOGAF™ Version 9*, Basharat Hussain.

Hauder, M., Matthes, F. & Roth, S., 2012. Challenges for automated enterprise architecture documentation. In *Trends in Enterprise Architecture Research and Practice-Driven Research on Enterprise Transformation*. Springer, pp. 21–39.

Huber, M. et al., 2017. The Influence of Shadow IT Systems on Enterprise Architecture Management Concerns. In *European, Mediterranean, and Middle Eastern Conference on Information Systems*. Springer, pp. 461–477.

IBM, 1978. *Business System Planning: Information System Planning Guide*,

Jugel, D., Sandkuhl, K. & Zimmermann, A., 2016. Visual analytics in Enterprise Architecture Management: a systematic literature review. In *International Conference on Business Information Systems*. Springer, pp. 99–110.

Kim, T.-Y. et al., 2006. A modeling framework for agile and interoperable virtual enterprises. *Computers in Industry*, 57(3), pp.204–217. Available at: https://search.proquest.com/docview/235768964?accountid=8144.

Korhonen, J.J. et al., 2016. Adaptive Enterprise Architecture for the Future: Towards a Reconceptualization of EA. *Informatics (CBI), 2016 …*, pp.272–281. Available at: http://ieeexplore.ieee.org/abstract/document/7780323/ [Accessed April 11, 2017].

Kotusev, S., 2017. Conceptual Model of Enterprise Architecture Management. *International Journal of Cooperative Information Systems*.

Kotusev, S., 2016. The History of Enterprise Architecture: An Evidence-Based Review. *Journal of Enterprise Architecture*, 12(1), pp.29–37.

KPMG, 2016. *Now or never - 2016 Global CEO Outlook*, Available at: https://home.kpmg.com/content/dam/kpmg/pdf/2016/07/ch-2016-global-ceo-outlook.pdf.

Lange, M., Mendling, J. & Recker, J., 2016. An empirical analysis of the factors and measures of Enterprise Architecture Management success. *European Journal of Information Systems*, 25(5), pp.411–431. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85016276209&doi=10.1057%2Fejis.2014.39&partnerID=40&md5=a115806775e77eee39796a426d1ec930.

Lapalme, J. et al., 2016. Exploring the future of enterprise architecture: A Zachman perspective. *Computers in Industry*, 79, pp.103–113. Available at: http://dx.doi.org/10.1016/j.compind.2015.06.010.

Lau, J.H. et al., 2011. Automatic labelling of topic models. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, pp. 1536–1545.

Nikpay, F., Ahmad, R.B., et al., 2017. An effective Enterprise Architecture Implementation Methodology. *Information Systems and e-Business Management*, pp.1–36. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85009446528&doi=10.1007%2Fs10257-016-0336-5&partnerID=40&md5=aa76b6715ccce7d27922eb974899a2e7.

Nikpay, F., Ahmad, R. & Yin Kia, C., 2017. A hybrid method for evaluating enterprise architecture implementation. *Evaluation and Program Planning*, 60, pp.1–16. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84988473835&doi=10.1016%2Fj.evalprogplan.2016.09.001&partnerID=40&md5=39a635a5e1ff07077303805df3387f63.

Nugraha, D.C.A. et al., 2017. A business architecture modeling methodology to support the integration of primary health care: Implementation of primary health care in Indonesia. *Internetworking Indonesia Journal*, 9(1), pp.39–45. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85017347432&partnerID=40&md5=37629637e68808569da5e3eca12b63c7.

Petersen, K., Vakkalanka, S. & Kuzniarz, L., 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, pp.1–18. Available at: http://dx.doi.org/10.1016/j.infsof.2015.03.007.

Petticrew, M. & Roberts, H., 2008. *Systematic reviews in the social sciences: A practical guide*, John Wiley & Sons.

Pinto, V.A., de Rezende Rohlfs, C.L. & Parreiras, F.S., 2014. *Applications of Ontologies in Enterprise Modelling: A Systematic Mapping Study*,

QualiWare, 2017. QualiWare Enterprise Architecture. Center of Excellence. Available at: https://coe.qualiware.com/ea-topic-model/.

Rahimi, F., Gøtze, J. & Møller, C., 2017. Enterprise architecture management: Toward a taxonomy of applications. *Communications of the Association for Information Systems*, 40(1).

Ramage, D. & Rosen, E., 2011. Stanford topic modeling toolbox.

Rouhani, B.D. et al., 2015. A systematic literature review on Enterprise Architecture Implementation Methodologies. *Information and Software Technology*, 62, p.1. Available at: https://search.proquest.com/docview/1668188955?accountid=8144.

Saint-Louis, P. & Lapalme, J., 2016. Investigation of the Lack of Common Understanding in the Discipline of Enterprise Architecture : A Systematic Mapping Study. *Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOCW*, 2016–Septe, pp.74–82.

Schneider, A.W., Schulz, C. & Matthes, F., 2013. Goals in Enterprise Architecture Management--Findings from Literature and Future Research Directions. In *Business informatics (cbi), 2013 ieee 15th conference on*. IEEE, pp. 284–291.

Simon, D., Fischbach, K. & Schoder, D., 2013. An Exploration of Enterprise Architecture Research. *Communications of the Association for Information Systems*, 32(1), pp.1–71. Available at: http://aisel.aisnet.org/cais/vol32/iss1/1.

Spewak, S.H. & Hill, S.C., 1993. *Enterprise architecture planning: developing a blueprint for data, applications and technology*, QED Information Sciences, Inc.

Turney, P.D. & Pantel, P., 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37, pp.141–188.

Wißotzki, M. & Sonnenberger, A., 2012. Enterprise architecture management-state of research analysis & a comparison of selected approaches. *actice of*, p.37.

Zachman, J.A., 1987. A framework for information systems architecture. *IBM systems journal*, 26(3), pp.276–292.

APPENDIX

Topic (number & label)	Coverage of the body of literature	5 most common words (word instances per topic)	Present in the content analysis of Simon et at. (2013)	Present in topics of Saint-Louis & Lapalme (2016)
2 Business Process	0,60%	process (56,63), basis (12,57), BusinessProcess (8,37), financial (6,81), integrate (5,12)	Included in the Business Architecture Layer	No
11 Framework	0,75%	framework (117,89), EnterpriseArchitectureFramework (8,12), integrated (7,91), structured (3,95), adequate (2,99)	No	EA-Framework
29 ERP Model	0,74%	model (17,80), erp (12,69), develop (7,20), companies (6,36), EnterpriseResourcePlanning (6,17)	No	No
52 Management	0,68%	management (51,17), managers (14,90), InformationSecurity (9,61), activities (8,79), information (7,15)	EA Management	Related to EA-Principles
102 Design	0,64%	design (31,78), level (10,63), BusinessProcesses (10,82), systems (9,75), component (9,71)	Similar to the Documentation phase	EA-Design
118 Modelling Methodology	0,62%	modeling (58,62), HealthCare (17,55), methodology (13,13), ModelingApproaches (4,85), rapidly (4,60)	Modelling element of the Methodology	Related to EA-Principles and EA-Strategy
211 Analysis Method	0,61%	method (55,63), analysis (26,92), CaseStudy (5,66), estimates (3,87), technique (3,44)	Analysis phase	Related to EA-Principles and EA-Strategy
227 Meta Model	0,78%	metamodel (20,41), analysis (13,99), models (8,53), perform (7,31), metamodels (6,81)	Meta Model element of the Methodology	No
319 Model	0,76%	model (128,19), integrated (5,95), evolution (3,14), contributes (2,96), systems (2,28)	Layers	EA-Model
355 Service	0,69%	service (43,66), services (39,65), resources (7,23), user (6,36), infrastructure (4,86)	No	No

Stopword list: enterprise, architecture, set, advances, number, presents, difficult, research, better, using, study, significant, important, conducted, paper, findings, contribute, approach, common, general, source, right, based, attempts, diverse, large, way, ensure, proposed, presented, web, proposes, allow, events, main, require, core, related, created, area, implemented, emerged, numerous, use, private, defines, cover, scope, provided, recent, years, various, elements, includes, specific, ea, aim, respect, different, multiple, small, nowadays, line, focuses, defined, major, new, increasing, insights, high, growing, helps, poor, issues, gap, just, existing, result, simple, make, approaches, extend, results, base, studies, key, used, step, article, matter, order, meet, enterprises, does, case, available, enable, business, review, future, provides, importance, means, needs, time, today's, field, initiatives, network, needed, need, group, called, open, uses, best, lack, organization's, e.g, useful, enterprise's, actual, making, work, setting, test, known, fact, typically, quickly, intended, shown, force, including, shows, primary, active, describes, fully, bring, increase, allows, enables, era, takes, offer, offers, presenting, calls, range, example, today, real, works, propose, consider, previous, cope, requires, focus, novel, suggested, introduced, establish, formal, highly, terms, makes, gives, joint, early, gain, viewed, development, help, contribution, question, questions, contributions, utilised, utilized, second, broad, emphasis, papers, suggests, concluded, described, concludes, plays, discusses, huge, continues, play, far, easy, given, applied, follow, ongoing, ii, manner, so-called, issue, ad, hand, iii, i, later, taken, exist, like, versus, old, run, lot, starts, box, vast, end, short, comes, going, left, author, authors, et, performance, lies, little, seen, ways, whilst, and/or, is/it